

To,
Product and Business Teams at Fetch

Subject: Data Modeling and Quality Insights for Fetch Data.

Hi Team,

I hope you are all doing well. I just performed an in-depth analysis using Python and SQL tools to extract and process Fetch data effectively, that revealed some major areas where our data quality might be enhanced. These enhancements will enable more informed business decisions and more accurate reporting.

Key Challenges and Business Impact:

Duplicate and Missing User Data

- **Issue:**
 - Over 50% of user records are duplicates, leading to inflated user counts.
 - Some user roles are labeled as "fetch-staff" instead of our default "consumer."
 - Key fields (e.g., sign-up source, state, and last login) are missing many records.
- **Business Impact:**
 - Misleading user analytics may affect strategies for customer growth and retention.
 - Duplicate records can skew marketing insights.

Inconsistencies in Brand Data

- **Issue:**
 - Brands table is structured at the brand level, but the barcode within it corresponds to individual items, creating a mismatch in granularity and leading to inconsistency.
 - Critical fields (brandCode, category, topBrand) have missing or inconsistent values.
 - Duplicate brand codes with conflicting attributes are causing unreliable reporting.
 - Presence of test data raises the question of whether it should be included in production analysis.
- **Business Impact:**
 - Incorrect brand categorization may lead to misleading insights and affect product recommendations.

Fragmented Receipt Data and Brand Mapping Issues

- **Issue:**
 - There is no well-defined relationship between receipts, receipt items, and brands.
 - Many records have missing totalSpent or purchasedItemCount values which leads to blank values for bonusPointsEarned and pointsEarned values.
 - Numerous barcodes are not mapped to brands.

- **Business Impact:**
 - Inaccurate revenue tracking and consumer behavior analysis.
 - Challenges in evaluating the effectiveness of loyalty programs and promotions.

Key Questions for Stakeholder Input

- **User Data:** Should “fetch-staff” users be included in our consumer analytics?
- **Brand Data:** Should we retain the test data in the brands table, or filter it out from production reports?

Proposed Improvement

- **Organize Data for Clarity:** Restructure how we store receipt and brand details, so every sale is clearly linked to its brand, making revenue tracking more reliable.
- **Enhance Data Entry:** Introduce simple checks at the point of data entry to catch errors, reduce duplicates, and fill in missing information.
- **Standardize Key Information:** Ensure consistent naming for user roles and brand identifiers. This clarity will improve the accuracy of our reports.

I'd love to discuss these points further and align on next steps. Please let me know your thoughts or if you have any additional input.

Best regards,
Rushabh Mehrotra